doi:10.1006/jmbi.2000.3951 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 301, 117–128

JMB

2.8 Å Crystal Structure of the Malachite Green Aptamer

Christopher Baugh†, Dilârâ Grate† and Charles Wilson*

Department of Biology and Center for Molecular Biology of RNA, Sinsheimer Laboratories University of California at Santa Cruz, Santa Cruz CA 95064, USA

*Corresponding author

Previous *in vitro* selection experiments identified an RNA aptamer that recognizes the chromophore malachite green (MG) with a high level of affinity, and which undergoes site-specific cleavage following laser irradiation. To understand the mechanism by which this RNA folds to recognize specifically its ligand and the structural basis for chromophore-assisted laser inactivation, we have determined the 2.8 Å crystal structure of the aptamer bound to tetramethylrosamine (TMR), a high-affinity MG analog. The ligand-binding site is defined by an asymmetric internal loop, flanked by a pair of helices. A U-turn and several non-canonical base interactions stabilize the folding of loop nucleotides around the TMR. The aptamer utilizes several tiers of stacked nucleotides arranged in pairs, triples, and a novel base quadruple to effectively encapsulate the ligand. Even in the absence of specific stabilizing hydrogen bonds, discrimination between related fluorophores and chromophores is possible due to tight packing in the RNA binding pocket, which severely limits the size and shape of recognized ligands. The site of laser-induced cleavage lies relatively far from the bound TMR (~15 Å). The unusual backbone conformation of the cleavage site nucleotide and its high level of solvent accessibility may combine to allow preferential reaction with freely diffusing hydroxyl radicals generated at the bound ligand. Several observations, however, favor alternative mechanisms for cleavage, such as conformational changes in the aptamer or long-range electron transfer between the bound ligand and the cleavage site nucleotide.

Keywords: in vitro selection; tetramethylrosamine; base quadruple; chromophore-assisted laser inactivation

Introduction

The relationship between RNA structure and function is complex; our understanding is hampered in part by the relatively small number of tertiary structures characterized to date. *In vitro* selection from pools of random sequence nucleic acids has yielded a variety of ligand-binding RNAs (''aptamers'') that specifically recognize a wide range of small molecules (Ellington & Szostak, 1990; Feigon *et al*., 1996; Hamasaki *et al*., 1998). The means by which several of these aptamers recognize and bind their ligands has been well studied by a combination of multidimensional NMR spectroscopy and X-ray crystallography (Dieckmann *et al*., 1996; Fan *et al*., 1996; Geiger *et al*., 1996; Jiang & Patel, 1998; Nix *et al*., 2000; Sussman *et al*., 2000; Yang *et al*., 1996; Zimmermann *et al*., 1997). For many of these RNAs, ligand induces a folding transition from an unstructured, unbound state to a highly ordered, bound state. This ''adaptive binding'' process mimics that also observed for a handful of well-characterized protein-binding RNAs, suggesting that it may represent a general property of RNA complexes (Patel *et al*., 1997). Because the energetics of ligand binding are interwoven with those of RNA folding, adaptive binding makes it particularly hard to predict the factors that will enable the formation of tight complexes. This problem, together with the relative paucity of structural information on functional RNA complexes, has hindered efforts at the rational design of RNA complexes (e.g. drug design against RNA targets).

The malachite green (MG) aptamer represents a potentially powerful model system for understanding the factors that stabilize RNA structure and

†Authors contributed equally to this work.

Abbreviations used: MG, malachite green; CALI, chromophore-assisted laser inactivation; TMR, tetramethylrosamine; ATP, adenosine triphosphate; FMN, flavin mononucleotide.

E-mail address of the corresponding author: wilson@biology.ucsc.edu

0022-2836/00/010117–12 $35.00/0

drive molecular recognition between RNAs and small molecules. One of a large family of commercial triphenyl methane-based dyes, MG has the unique property of generating destructive radicals when irradiated with high-intensity red laser light (Liao *et al.*, 1994). Using MG-tagged antibodies, several groups have exploited this property to target the *in situ* destruction of specific protein antigens (Jay & Keshishian, 1990; Jay & Sakurai, 1999; Liao *et al.*, 1994; Wang *et al.*, 1996). With the impetus of enabling a similar method for knocking out RNA molecules, we previously isolated a aptamers that binds MG (Grate & Wilson, 1999). Eight cycles of selection on MG-derivatized agarose followed by enzymatic amplification yielded a collection of ~40 independent sequences capable of specific MG recognition. Functional sequences did not appear to contain a common functional motif, although many carried a short conserved sequence of six nucleotides (CGAAUG). The characterization of one of these clones by truncation and mutagenesis suggested that binding could be ascribed to an internal loop containing the conserved hexanucleotide and flanked on either side by RNA helices. Transcripts containing this loop in the proper context can be targeted for laser-mediated destruction (Grate & Wilson, 1999). While antibody-directed ablation of protein targets is generally believed to involve non-specific free radical attack at many sites, laser-irradiation of the MG aptamer induces highly localized cleavage of the RNA backbone, suggesting that a specific mechanism for free radical destruction might operate (Grate & Wilson, 1999).

As with the adenosine triphosphate (ATP) and flavin mononucleotide (FMN) aptamers, the secondary structure of the MG aptamer consists of a conserved internal loop flanked by RNA helices (Burgstaller & Famulok, 1994; Sassanfar & Szostak, 1993). The NMR structures of ATP, theophylline, and FMN-binding RNAs have shown that the aromatic ligands are bound in active sites defined largely by stacking interactions built with unusual base interactions and/or backbone conformations (Dieckmann *et al.*, 1996; Fan *et al.*, 1996; Jiang *et al.*, 1996; Zimmermann *et al.*, 1997). Specificity is provided in both cases by additional hydrogen bonding between one or more face of the ligand and binding pocket nucleotides in a manner reminiscent of conventional base-pairing. The ATP and FMN aptamers have provided a wealth of information on the ways in which nucleotides can interact in non-canonical arrangements to stabilize tertiary structure, highlighting in particular the role of purine-purine base-pairs, and on the ways RNA backbone conformations can deviate dramatically from standard *A*-form geometry (Feigon *et al.*, 1996; Hermann & Patel, 1999).

To understand further the requirements for the formation of specific, stable complexes between small molecules and RNAs, we have determined the 2.8 Å crystal structure of the MG aptamer complexed with tetramethylrosamine, a high-affinity ligand. As with the ATP and FMN aptamers, numerous non-conventional base-pairings stabilize a tertiary structure that folds around the bound ligand. Remarkably, no hydrogen bonds are formed between the RNA and its ligand. Instead, specific binding appears driven largely by electrostatic and steric complementarity.

Results

While we were able to crystallize the MG aptamer complexed with malachite green under a variety of different conditions, diffraction from these crystals never extended beyond 5.5 Å. To obtain crystals that would be useful for structure determination, we tested a variety of modified complexes in which either flanking helices or the hairpin loop were altered, or in which the malachite green ligand was substituted for analogs. The fluorophore tetramethylrosamine differs from malachite green by the addition of a bridging oxygen ion that connects the two dimethlyamine-bearing phenyl rings and locks them into a planar orientation. The aptamer's affinity for TMR is substantially stronger ($K_D \sim 40$ nM) than that for malachite green ($K_D \sim 800$ nM), and thus we attempted crystallization using this alternative ligand. The TMR-aptamer complex crystallized in a habit similar to the MG-containing crystals, and the diffraction limit improved to 2.8 Å.

Crystallographic phases were determined by multiple wavelength anomalous dispersion using the scattering from bromines incorporated into the RNA through *in vitro* transcription. By replacing UTP in the standard transcription reaction with 5-Br-UTP, all five uridine bases in the sequence could be substituted with 100 % efficiency. The brominated form of the aptamer binds TMR with marginally higher, affinity ($K_D \sim 20$ nM) than the native form, and also proved suitable for crystallization, yielding crystals that diffract to 2.8 Å. A five-wavelength MAD dataset from a single crystal was used to determine crystallographic phases and solve the structure as described in Materials and Methods. The structure is currently refined to an *R*-factor of 26.094 % and R_{free} value of 28.22 % (Table 1).

5-Bromouridine has been used extensively to obtain phasing information for nucleic acid crystals (Anderson *et al.*, 1999; Correll *et al.*, 1997; Scott *et al.*, 1995; Shepard *et al.*, 1998; Todd *et al.*, 1999; Wild *et al.*, 1999). For the most part, however, it has been incorporated by chemical rather than enzymatic synthesis. Enzymatic synthesis is technically much simpler and provides significantly higher yields. Given the ability of automated Patterson peak searching algorithms to interpret accurately scattering from crystals containing many independent heavy atom sites (Brünger *et al.*, 1998; Huber *et al.*, 1997; Terwilliger & Berendzen, 1999), the strategy used here seems likely to be of general use for many small-to-medium-sized

Table 1. X-ray data collection, MAD phasing, and structure refinement

Data set	λ (Å)	d_{min} (Å)	Observed reflections[a]	Unique reflections[b]	%Complete	$\langle I\rangle/\langle\sigma_i\rangle$	R_{sym} (%)[c]	Centric ⟨PP⟩[d]	Acentric ⟨PP⟩[d]	
								Iso	Iso	Anom
Low remote	1.00000	2.8	124,894	2946	91.1	30.36	5.8	-	-	2.86
Inflection point	0.92090	2.8	116,373	2930	91.1	26.02	7.0	2.08	3.12	1.00
Peak	0.92070	2.8	112,230	2935	91.1	28.79	6.5	1.58	2.51	3.60
High remote	0.90830	2.8	105,438	2936	91.1	28.30	7.2	2.26	4.00	3.16
Second peak	0.92070	2.8	128,684	3188	99.7	40.59	5.7			
Figures of merit	MAD	0.70		*R*-factor[e]	R_{cryst} (%)	26.09		RMSD[f] from ideality	Bond lengths (Å)	0.013
	Solvent flattened	0.91			R_{free} (%)	28.22			Bond angles (deg.)	1.96

[a] Reflections indexed out to 2.5 Å resolution to ensure appropriate orientation matrix calculation.
[b] Reflections scaled out to 2.8 Å due to a drop off in $\langle I\rangle/\langle\sigma_i\rangle$.
[c] $R_{sym} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\langle I(h)\rangle$, where $I_i(h)$ is the Ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
[d] Phasing power is the mean value of the heavy atom structure factor amplitude divided by the residual lack of isomorphous (iso) and anomalous (anom) closure error for centric and acentric reflections.
[e] *R*-factor $= \Sigma|F_o - F_c|/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is the cross-validated *R*-factor calculated for 10% of the reflections omitted in the refinement process.
[f] EMSD is the root-mean-square deviation.

RNAs in which bromo-substitution does not interfere with folding.

Structure overview

The secondary structure of the aptamer, defined by two Watson-Crick helices flanking an asymmetric loop, matches that predicted on the basis of previous deletion and mutagenesis experiments (D.G., C.B., unpublished observations). Folding of the asymmetric loop is stabilized by a variety of stacking and base-pairing interactions that together combine effectively to encapsulate the ligand (Figure 1). With one exception (U32), all of the internal loop nucleotides are directly involved in the formation of the binding pocket. Base stacking within the loop results in the formation of a pseudo-continuous helix connecting the flanking stems. The helix 2 hairpin is closed by an engineered GNRA-type tetraloop (nucleotides 15-18) with the sequence 5′-GAGA-3′. The structure of this loop resembles that for other GNRA tetraloops with the exception of the second nucleotide, A16. In all other loops of this type whose structures have been determined, the ''N'' nucleotide is stacked upon the following purine (''R'') base. In the MG aptamer structure, however, this nucleotide is flipped over onto the preceding conserved G of the tetraloop. Given that the flipped adenosine residue allows a crystal contact to form between the following guanosine and a stacked uridine nucleotide from a symmetry related molecule (Figure 2), we presume that this altered conformation may exist only in the context of the crystal lattice. With the exception of the first two base-pairs in helix 2, the helices and the hairpin loop appear to play no special role in building the aptamer active site other than to bring the two halves of the internal loop into contact with each other. Replacement of base-pairs in helix 1 or base-pair proximal to the closing tetra-loop in helix 2 yields molecules as active as the wild type sequence (D.G., C.B., unpublished observations). Analysis of the *B*-factors in the two stems and the binding pocket indicates that helix 1 is the least constrained ($\langle B\rangle = 47$ Å^2) with somewhat better crystallographic order in helix 2 and the bulge region ($\langle B\rangle = 39.5$ Å^2 and 42.4 Å^2, respectively). The ligand itself represents one of the most highly constrained residues in the structure ($\langle B\rangle = 22.5$ Å^2), consistent with its positioning deep within the aptamer complex.

Strontium is absolutely required for crystallization of the bromine-substituted RNA (other divalents, such as calcium or barium, will not substitute for it). This requirement can be readily explained in part by the presence of a specifically bound strontium ion that bridges between symmetry related molecules in the crystal lattice. The details of the metal binding site together with representative electron density are shown in Figure 2. The strontium ion is coordinated by the 02 of U32 and by a non-bridging phosphate group oxygen atom of the following nucleotide (G33), effectively locking the uridine molecule up and away from the loop. In addition, the strontium ion contacts a phosphate oxygen atom from the A16 of a neighboring molecule (the 'misplaced' tetraloop ''N'' nucleotide). The limited resolution of the structure makes it impossible to determine the exact coordination state of this metal, although the range of distances between the metal and the RNA oxygen atoms (2.6-2.9 Å) suggest direct, inner sphere contacts rather than indirect, water-

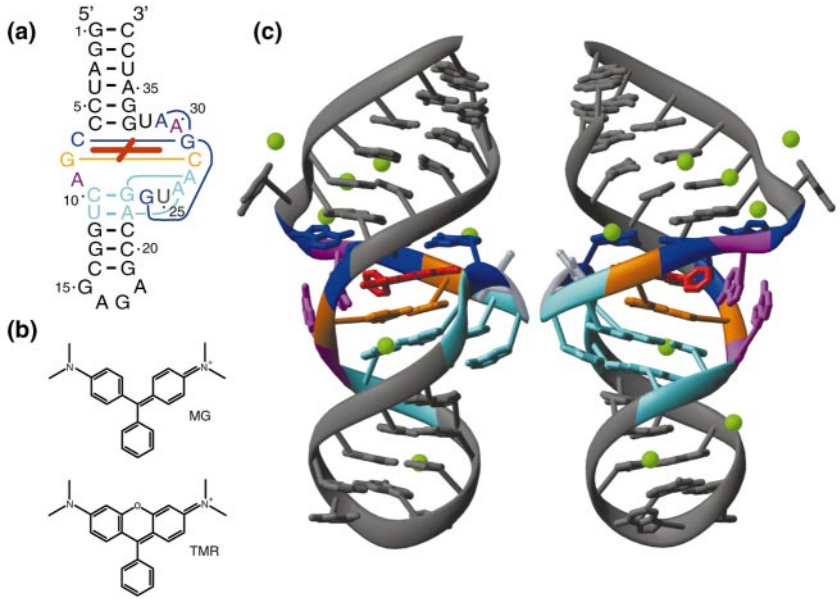

Figure 1. (a) Secondary structure schematic of the aptamer. (b) Chemical structure of malachite green (MG) and tetramethylrosamine (TMR). The aptamer was selected to bind MG and the structure was solved with TMR as the ligand. (c) Front and back schematic views of the tertiary structure. Color coding for both the secondary and tertiary structure schematics is as follows: TMR, red; the canonical G8:C28 base-pair, orange; U25, the uridine involved in the U-turn, silver; the A26·U11:A22 and A27·C10:G23 base triples, cyan; the G24·A31·G29:C7 base quadruple, blue; the perpendicularly stacking A9 and A30 adenosines, magenta; bound strontium ions, green. Figures prepared using RIBBONS (Carson & Bugg, 1986).

mediated contacts (Mosges *et al.*, 1992; Mueller *et al.*, 1999). Seven additional strontium ions are also observed in the structure. In contrast to the U32 site, however, these ions appear to be less well ordered in the crystal and indirectly to coordinate the major groove face of guanosine nucleotides in the helical regions of the aptamer. No metal ions are found in the core of the aptamer, consistent with the observation that activity is maintained in the absence of all divalent metals (D.G. & C.B., unpublished observations).

Tier structure within the asymmetric loop

The binding pocket of the aptamer can be considered as a series of stacked nucleotide tiers that include one base quadruple, two minor groove base triples and one conventional base-pair. Oriented perpendicularly to these layers are two additional nucleotides that stack upon themselves and the ligand. The ligand is intercalated within these tiers, and sits nearly completely surrounded in the binding pocket. At the center of each tier lies a conventional Watson-Crick base-pair that is (in all but one case), elaborated upon by further minor and/or major groove base interactions.

Canonical G:C base-pairs (C7:G29 and G8:C28) stack directly above and below the xanthene ring of the ligand to form the center of the binding pocket. Insertion of TMR between C7 and G8 on one side of the loop, and between C28 and G29 on the other, is accomplished without any significant distortion of the base-pairing. The C7:G29 pairing on the helix 1 side of the ligand is augmented by nucleotides G24 and A31 that contact the pair's major and minor groove faces to create a novel base quadruple (discussed below). In contrast, the G8:C28 pairing on the helix 2 side of the ligand exists as an isolated Watson-Crick base-pair. Below it, stacked adenosine residues from the long strand of the internal loop lie in the minor groove of helix 2. Together, they form a set of base triples that effectively pin the lower portion of the aptamer core in place. Connecting nucleotides (A9, U25, A30) link the tiers above and below the ligand to complete the binding pocket.

In the base triple most distant from the ligand, A26 docks in the minor groove of the U11:A22 base-pair. The N1 and N6 of the loop adenosine are hydrogen bonded respectively to the 2′-hydroxyl and O2 of U11. Lying between this triple and the ligand, a second minor groove triple is formed by A27, C10, and G23. Whereas A26 interacts predominantly with the 5′-strand of helix 2, A27 is aligned with the 3′-strand and its N1 and N6 hydrogen bonded to the N2 and N3 of G23. Details of the hydrogen bonding in these base triple interactions are shown in Figure 3(a) and (b).

Minor groove A·X:Y base triples (X:Y = a Watson-Crick base-pair) have been observed in several structures and may represent one of the more common tertiary interations that stabilize RNA structure. One of the surprising features of this interaction is its relative plasticity. Adenosine has been observed to pair with all four types of nucleotides in the minor groove and considerable variation in hydrogen bonding patterns exist for base triples with the same composition. In the MG aptamer, the beet western yellow virus frameshifting pseudoknot, and the biotin aptamer, docking between consecutive loop adenosine bases and base-pairs in a helix produces a series of stacked triples (Nix *et al.*, 2000; Su *et al.*, 1999). In all of these cases, the highly favored adenosine-adenosine stacking is likely to play an important role in defining their energetic contribution to RNA fold-

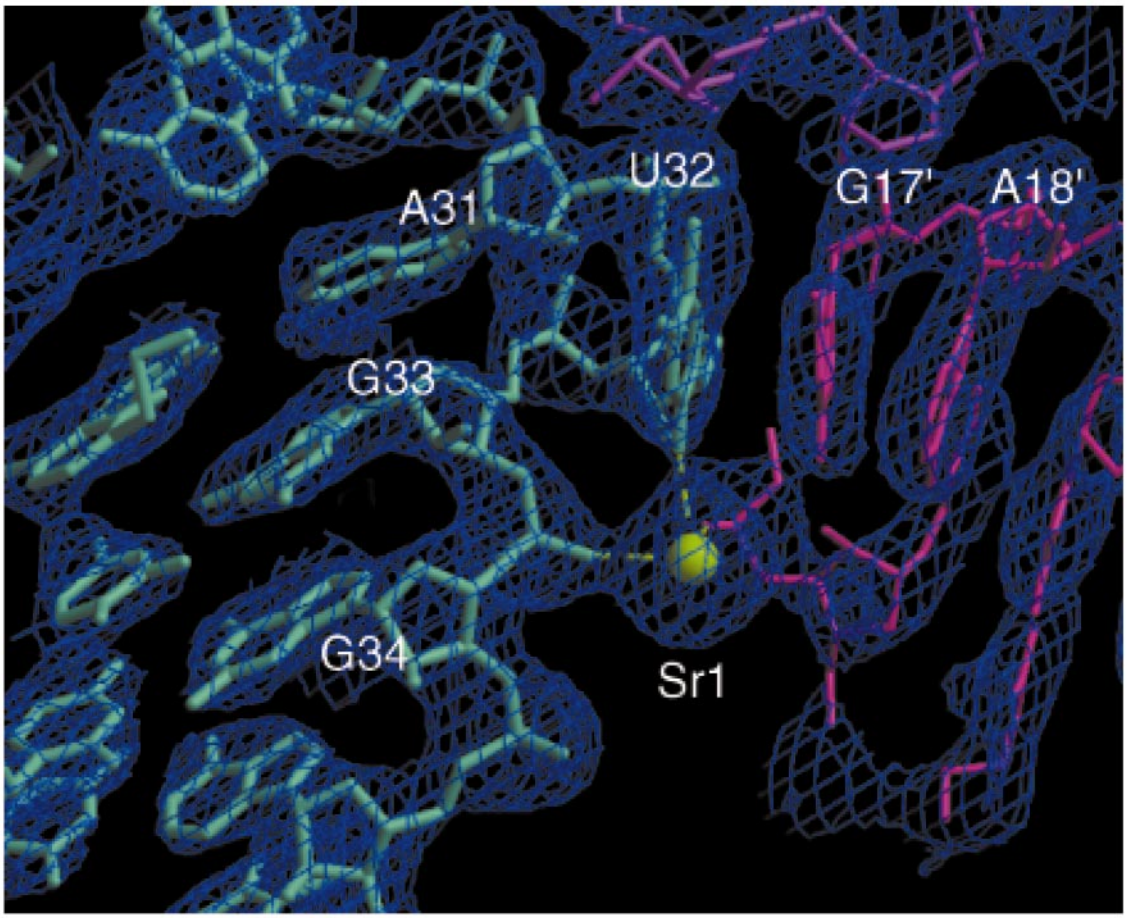

Figure 2. Strontium ion-binding site. σ_A-weighted $2F_o - F_c$ electron density map of the coordinated strontium. Sr1 is directly coordinated by the O2 of U32 and the O2P of G34 in the aptamer in cyan, and also by the O2P of A16′ of a symmetry-related molecule in pink. A16′ is flipped out of the canonical GNRA tetra loop position to allow the direct stacking of U32 with G17′. Figure generated by Raster3D (Merritt & Murphy, 1994).

ing and may partly explain the phylogenetic preference for adenosine bases in the loop sequences.

Completing formation of the stacked layers in the aptamer core is a novel, in-plane base quadruple (Figure 3(c)). The quadruple can be considered as the union of a conventional adenosine minor groove triple (virtually identical to that formed by A27·C10:G23) and a conventional guanosine major groove triple (matching that formed by C13:G22·G46 in $tRNA^{Phe}$) (Stout *et al.*, 1976). The net effect is to surround completely G29 on every face and thereby create a large (~230 Å^2) platform upon which the ligand sits. A similar arrangement between nucleotides and a planar chromophore has been observed in the FMN aptamer complex (Fan *et al.*, 1996). In this particular case, a somewhat smaller base triple formed in an internal loop defines the platform. A handful of base quadruples have been reported in previous nucleic acid structures. Intermolecular G-quartets, formed by cyclically arranged G·G pairs, have been observed in both telomeric sequence DNAs and in an RNA quadruplex (Katahira *et al.*, 1995; Williamson, 1994). In both the beet western yellow virus pseudoknot and the hepatitis delta virus ribozyme, a protonated cytidine molecule docks on the major grove face of a helix to form bifurcating hydrogen bonds to a pair of stacked nucleotides (Ferré-D'Amaré & Doudna, 2000; Su *et al.*, 1999). Conventional Watson-Crick base-pairing by one of these nucleotides yields a set of four bases linked by hydrogen bonds, i.e. a base quadruple. The arrangement in the MG aptamer is fundamentally different from these previous examples, in that a single nucleotide pairs directly with the three other nucleotides in the quadruple. The quadruple interaction in the MG binder serves to bring together four separate corners of the asymmetric loop, and thus define the overall fold of the aptamer.

Binding pocket architecture

The backbone reverses direction several times in the core as it tracks above and below the ligand. Three nucleotides (U25, A9, and A30) not directly involved in base-pairing are positioned to facilitate these turns and to bring together the tiers around the TMR. U25 links G24 in the base quadruple

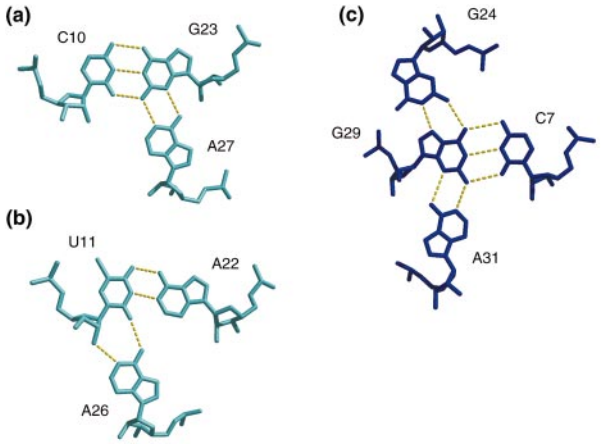

Figure 3. Hydrogen bonding for tiers within the aptamer core. All views are looking down from stem 1 towards stem 2. (a) In the triple interaction most proximal to the loop, A27 docks on the 3′-side of the minor groove of helix 2. (b) Stacked below, A26 docks on the 5′-side of the minor groove of helix 2. (c) The core of the aptamer is defined by a base quadruple centered around G29. The minor groove, Watson-Crick, and major groove faces of G29 are paired with A31, C7, and G24 respectively. All Figures made with Conic (Huang *et al.*, 1991).

with A26, the lower base triple involving helix 2. Together U25, A26, A27, and C28 define a U-turn virtually identical to that originally discovered in $tRNA^{Phe}$ and later observed in the catalytic core of the hammerhead ribozyme and the U2 snRNA (Figure 4). In all four cases, the defining uridine of the U-turn forms a conserved pair of hydrogen bonds with the following nucleotides (Figure 4). U-turns provide a ready mechanism for reversing the path of an RNA chain, while simultaneously leaving the Watson Crick faces of following nucleotides in a conformation available for pairing (Jucker & Pardi, 1995). In the MG aptamer, the bases following the U-turn uridine are positioned to pair with helix 2 to form the minor groove adenosine triples described earlier.

The aptamer core is completed by A9 and A30, lying at opposite corners of the internal loop. The base of A30 stacks directly against the phenyl ring of TMR, while the A9 base is flipped upside down relative to A30 to pack against it (Figure 5). In addition to providing additional contacts with the ligand, this top-to-bottom interaction helps further bring together the two sides of the bulge, and hold them together in a tightly compact arrangement. Through this combination of pairing and stacking interactions, the aptamer is able virtually to encapsulate the ligand, leaving <20 % of its surface area exposed to solvent.

Laser-mediated cleavage

Previous studies had shown that laser irradiation of the malachite green-aptamer complex causes highly localized cleavage, suggesting a specific mechanism might serve to position MG-generated hydroxyl radicals directly for RNA attack (Grate & Wilson, 1999). The predominant cleavage site lies between the phosphate groups of A26 and A27 within the U-turn at a position surprisingly far from the bound ligand (Figure 4). Hydroxyl radicals generated by the Fenton reaction using metals chelated in solution (e.g. Fe(II)-EDTA, Cu(I)-phenanthroline) or tethered at specific sites within a macromolecule (e.g. Fe(II)-BABE) have been used extensively to probe the structures of RNA and DNA complexes (Culver & Noller, 1998; Heilek & Noller, 1996; Hermann & Heumann, 1995). Cleavage efficiency generally correlates well with

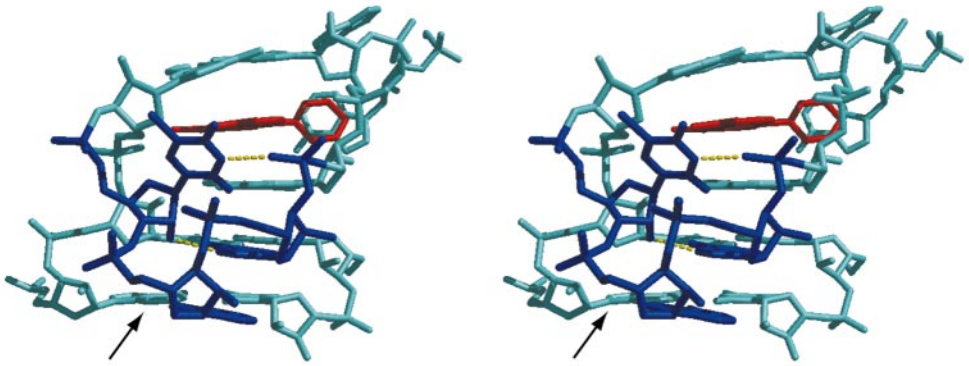

Figure 4. Stereo view of the U-turn and the cleavage site. Specific hydrogen bonds in the U-turn (blue) are shown as yellow lines. The bound ligand (red) and the cleavage site nucleotide, A26 (arrow), form the ends of a continuous aromatic stack. Figure created using Conic (Huang *et al.*, 1991).

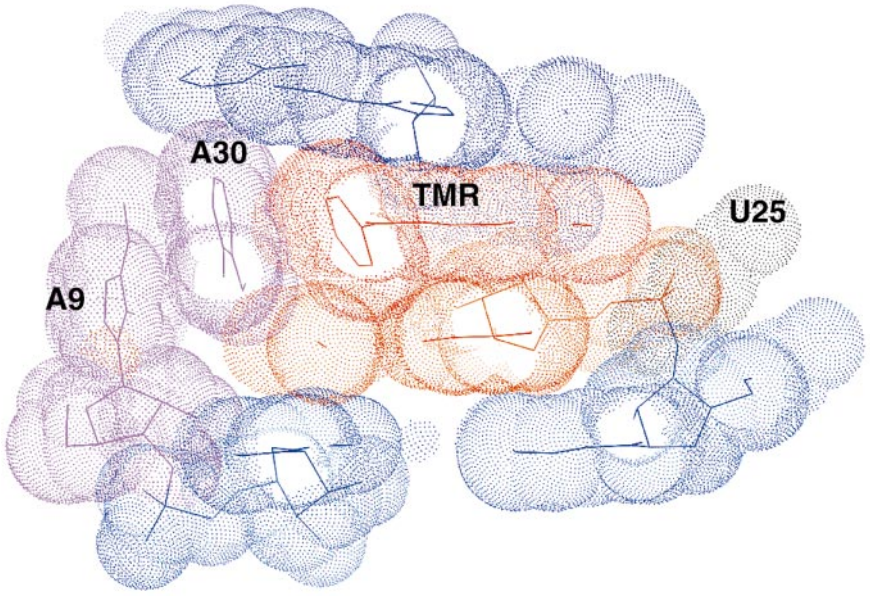

Figure 5. Details of the packing in the binding pocket. The base quadruple (blue) and the central base-pair (orange) sandwich above and below respectively to form most of the ligand contact surface. Stacking of A9, A30 (magenta) against the free phenyl ring of the TMR (red) and the packing of U25 (silver) against one of the dimethylamines of the TMR further delimit the sides of the binding pocket. Figure created with MIDAS (Ferrin *et al.*, 1988).

solvent accessibility at the ribose C4′ and C5′ positions (Balasubramanian *et al.*, 1998) and falls off rapidly as a function of distance when the radical generator is immobilized (Hermann & Heumann, 1995). If free radicals are in fact generated directly at the ligand, they must diffuse ~15 Å before reacting with A26. Preferential attack at this site could be explained if it were conformationally predisposed for reaction with freely diffusing hydroxyl radicals. While A26 adopts an unusual conformation, several lines of evidence argue against this. The accessibility of ribose atoms in the cleaved nucleotide is no higher than that for several other loop nucleotides (including G24, which directly contacts the ligand and appears better positioned for attack). Fe(II)-EDTA treatment of the MG aptamer complexed with either MG or TMR modifies A26 to the same extent as almost all other nucleotides (D.G., unpublished observations), indicating that the site is not intrinsically hyper-reactive to solution free radicals. Chemical modification studies using Cu(I)-phenanthroline similarly suggest that the equivalent position in the $tRNA^{Phe}$ U-turn (G_m34) is not significantly activated for attack (Hermann & Heumann, 1995). Cu(I)-phenanthroline is believed to intercalate between the bases of A35 and A36 and cleavage at neighboring nucleotides (including G_m34) falls off uniformly as a simple function of the predicted distance from the bound copper ion (Hermann & Heumann, 1995). Together, these observations suggest that laser-induced cleavage might involve a mechanism other than long-range diffusion of hydroxyl radicals.

It is possible that the aptamer transiently adopts a different conformation that positions A26 nearer to the bound ligand (although this would require a significant refolding of the aptamer and seems unlikely, given the constrained pairing arrangements within the aptamer core). Alternatively, free radicals might be generated directly at the cleavage site by a different mechanism. Studies using modified nucleotides and site-specific intercalators have suggested that electrons can tunnel through a DNA helix from one site to another through the stacked bases that link them (Kelley & Barton, 1999). We note that the cleavage site adenosine lies at the end of a continuous base stack that connects to the center of the buried TMR (Figure 4). Future experiments will test the possibility that electron transfer through these bases might serve to mediate laser-induced cleavage of the aptamer.

Binding site interactions between TMR and the aptamer

Jain & Sobell (1984) have characterized the interaction between the non-specific intercalating agent, ethidium, and a CpG dinucleotide, using X-ray crystallography. In the complexed form, the dinucleotide dimerizes to form G:C base-pairs that sandwich the bound dye. The entire ethidium complex (NDB identifier DRB006) can be superimposed remarkably well with the central G:C base-pairs of the MG aptamer and its bound ligand. Given the similarities between the two structures (Figure 6), we can proceed to ask how other elements in the aptamer core combine to add specificity to ligand

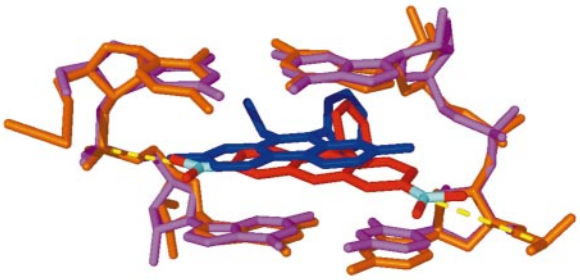

Figure 6. Comparison of the MG and ethidium binding sites. The CpG dimer-ethidium ternary complex (magenta-blue; NDB identifier DRB006) is superimposed with the central C7:G29/G8:C28 base-pairs and TMR of the aptamer complex (orange-red). Pairings between the positively charged dimethylamine groups of TMR and the nearest phosphate backbone oxygen atoms of the aptamer are shown as broken yellow lines. Figure created with MIDAS (Ferrin *et al.*, 1988).

binding. It is worth noting that although TMR, MG, and ethidium have some structural similarities, their intrinsic affinities for nucleic acids are fundamentally different. Ethidium has a high level of affinity for both RNA and DNA duplexes and intercalates at random sites largely irrespective of sequence (slightly preferring G:C base-pairs). In contrast, TMR and MG have minimal non-specific affinity for either. Less than 0.1% of random sequence ($N = 72$) RNA is retained on a malachite green column and $<10^{-12}$ bind with a high level of affinity (Grate & Wilson, 1999).

Interactions driving complex formation are dominated by planar stacking between the aromatic rings of TMR and the bases of the aptamer. In comparison to the ethidium complex, TMR is shifted to one side of the core base-pairs, leaving one of the dimethylamine groups and part of the xanthene ring protruding outwards from the major groove edge. Formation of the base quadruple positions G24 to maintain close contacts with this part of the ligand. In addition, the perpendicularly oriented face of the U25 base (locked in position by hydrogen bonding to the backbone) forms a pocket around one of the amine methyl groups, providing additional contacts that may stabilize binding (Figure 5). Both ethidium and TMR contain phenyl groups attached to the main tricyclic nucleus of each molecule. In the ethidium complex, this moiety protrudes from the minor groove face without making any direct RNA contacts. In contrast, the equivalent ring on TMR is sandwiched in the aptamer structure by the unpaired, stacked adenosine bases (A30, A9) on one face and by the G29 ribose on the other. The phenyl ring and the A30 base are tilted slightly with respect to each other, limiting the extent of their contacts. This imperfect packing may be a consequence of the constraints of the crystal lattice as we note that the A9 base makes a direct stacking contact with the equivalent base from a symmetry related molecule. As such, packing appears optimized for the intermolecular A30-A9-A9′-A30′ stack at the apparent cost of close contacts with the ligand.

In addition to the extensive steric complementarity noted above, electrostatic interactions between the cationic ligand (with a net +1 charge distributed between the two amine groups) and the anionic RNA contribute to binding. Methyl groups attached to the amine groups closely approach or contact the pro-R_P oxygen atoms of C8 and G28 (Figure 6), preventing their direct solvation and providing a low dielectric microenvironment for the favorable electrostatic interaction between the amine and the phosphate groups. Completely absent from the complex are hydrogen bonds between heteroatoms on the ligand and the RNA. As such, the mode of ligand recognition is fundamentally different to that observed in the ATP and FMN-aptamer complexes where base-ligand hydrogen bonding is critical for specificity. We note that FMN does not bind detectably to the MG aptamer, nor does the FMN aptamer bind TMR (data not shown).

A variety of ligand analogs are commercially available and their relative levels of binding affinity can be used to infer the energetic contribution of ligand-RNA interactions observed in the aptamer crystal structure (Figure 7). MG differs from TMR by the omission of the bridging oxygen ion in the xanthene ring (Figure 1(b)). Removal of this oxygen atom allows the dimethylaniline rings to rotate freely and reduces the conjugation within the ring system; effects expected to reduce the ligand-RNA stacking interaction energy. Consistent with this prediction, we find that the level of affinity for MG is reduced somewhat relative to that for TMR ($K_D \sim 800$ nM *versus* 40 nM), corresponding to a reduction in binding energy by 1.8 kcal/mol. Surprisingly, we are unable to detect binding by crystal violet (differing from malachite green simply by the addition of a dimethlyamine to the free phenyl ring; Figure 7). This result is especially remarkable given that the additional amine group is positioned such that it should not introduce any steric clashes or close contacts with the RNA. Two effects likely account for the inability to bind. Some of the net positive charge localized to the two amine groups on MG is distributed to the third amine group in crystal violet, reducing the favorable electrostatic interactions between the amine and the backbone phosphate groups. The additional charge delocalization also drives an altered arrangement of the dimethylaniline rings as shown by crystallographic studies of the isolated dyes (Figure 7). Whereas the aniline rings in MG are only slightly out-of-plane with respect to each other and the phenyl ring is roughly perpendicular to them, all three rings in crystal violet show the same significant propeller twist. If this conformation were maintained upon binding, significant

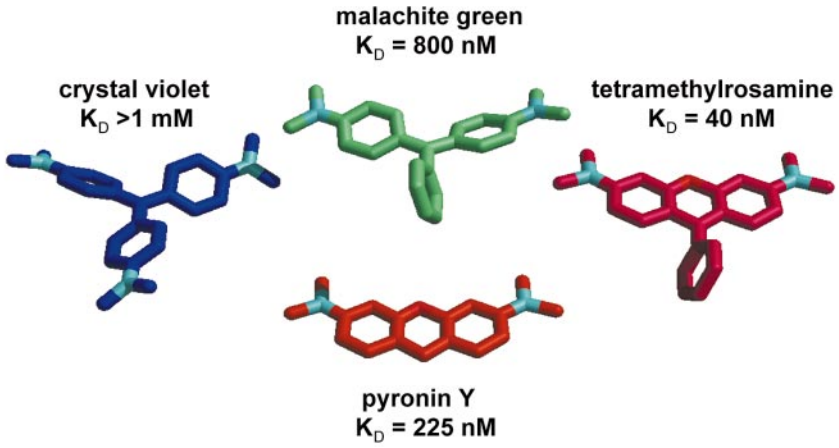

Figure 7. Binding by ligand analogs. Aptamer binding was determined as described in Materials and Methods for TMR, MG, pyronin Y, and crystal violet by measuring changes in fluorescence anisotropy. The structures shown for TMR, MG, and crystal violet were determined by X-ray crystallography (TMR in the current report; MG and crystal violet as isolated dyes, CCDC identifiers NACWOD and CEXDOY, respectively). The structure for pyronin Y was modeled based on that for TMR. As indicated, the levels of binding affinities for the different ligands correlate well with their relative planarity. Figure created with Conic (Huang *et al.*, 1991).

steric clashes with the aptamer binding site nucleotides would result. Comparing the observed crystal structures of TMR, MG, and crystal violet, we note a good correlation between binding affinity and the overall planarity of the conjugated rings in each.

Pyronine Y is identical to TMR, except that it lacks a free phenyl ring attached to the xanthene ring (Figure 7). Removal of the phenyl ring causes the binding affinity to drop ($K_D \sim 225$ nM), corresponding to free energy loss of ~1.0 kcal/mol. This relatively small cost in binding energy is surprising, given the extensive contacts that form between the phenyl ring and the RNA. Given the small effect, we presume that the conformation of A9 and A30 may be altered in the pyronine Y complex but that their mutual stacking is maintained.

Conclusion

In several respects, the TMR-aptamer complex resembles a globular protein with the dye serving as a hydrophobic core. Formation of a novel base quadruple is key to building a surrounding RNA scaffold that accommodates the large ligand. While the aptamer is clearly highly adapted to specifically recognize MG and its analogs, there is good reason to believe that similar small molecule binding sites might exist in biological RNAs. For example, the antibiotic tetracycline (which like TMR contains four rings and a dimethylamine) specifically binds and photocrosslinks within domain V of bacterial ribosomal RNA (Steiner *et al.*, 1988). By understanding in detail the energetics and structural requirements for small molecule binding to RNA, we will be better positioned to design *de novo* compounds that can bind and inhibit therapeutically relevant RNA targets.

Materials and Methods

The evolution and initial characterization of the aptamer has been described (Grate & Wilson, 1999). Aptamer RNA was transcribed from an EarI linearized pGEM-3Z plasmid containing the insert sequence 5′–CATATG-TAATACGACTCACTATAGGATCCCGACTGGCGAGA GCCAGGTAACGAATGGATCCAGAAGAGGAATTC-3′. Large scale *in vitro* transcriptions were performed in 40 ml of 20 mM $MgCl_2$, 40 mM Tris (pH 8.1), 0.1% (v/v) Triton X-100, 0.015% (w/v) spermidine, 5 mM DTT, 5 mM each of ATP, CTP, GTP, and 5-Br-UTP in the nucleotide mix with 50 μg/ml of template DNA and an optimized amount of T7 RNA polymerase and incubated overnight at 37°C. RNA was precipitated and desalted with a Sephadex G-50 column and subsequently purified on a preparative 20% PAGE gel. The RNA was excised from the gel and recovered by the crush and soak method in 0.3 M sodium acetate, ethanol precipitated, and resuspended in water.

Crystallization

Crystals were grown using the hanging drop vapor diffusion method. A solution containing 3.33% (v/v) MPD, 13.2 mM Na-Cacodylate (pH 4.5), 165 μM spermine, 23.16 mM KCl, 39.6 mM $SrCl_2$, 6.66 mM $MgCl_2$, 500 μM MGB RNA, 1 mM TMR was equilibrated against 35% MPD at room temperature. Crystals generally appeared after five days and grew to a maximum size of 200 μm × 200 μm × 100 μm over ten days. Crystals were prepared for data collection by rinsing in the drop conditions without RNA or ligand and with additional 35% MPD as a cryoprotectant for approximately ten minutes. Crystals were flash frozen in liquid propane and stored in liquid nitrogen.

X-ray data collection

A four wavelength MAD data set was collected on a single crystal at the Advanced Light Source beamline 5.0.2, using the bromine atoms from the incorporated 5-Br-uridine as the anomalous scattering source. Data from a frozen crystal were collected as a series of 1° oscillations with inverse beam geometry. Wavelengths for the MAD experiment were determined from an X-ray fluorescence scan of a 10 mM solution of 5-Br-uridine. A fifth data set was collected on the same crystal to maximize the completeness of the data. Reflections were indexed, integrated, and scaled using Denzo and Scalepack (Otwinowski & Minor, 1997). The crystals belong to the hexagonal space group $P6_522$, with unit cell parameters $a = b = 55.297$ Å, $c = 143.92$ Å. The asymmetric unit contains one aptamer-ligand complex.

Structure determination and refinement

Initial estimates for the positions of the bromine atoms in the unit cell were determined using an automatic Patterson search method implemented by the SOLVE program (Terwilliger & Berendzen, 1999). These sites were further refined and experimental phases were calculated with the program SHARP (De La Fortelle & Bricogne, 1997). Experimental phases were improved by density modification (using the SOLOMON program (Collaborative Computational Project, 1994)) to yield an easily interpretable map. Initial models built into the electron density were refined against the most complete fifth data set using CNS (version 0.9) (Brünger *et al.*, 1998). Statistics for the refinement are summarized in Table 1.

Measuring binding affinity

Binding of the fluorophores TMR and pyronine Y to the aptamer was measured by monitoring changes in fluorescence intensity or anisotropy using a Perkin Elmer LS50B Luminescence Spectrometer. Binding of the chromophores MG and crystal violet was inferred by competitive inhibition of TMR binding. Concentrations of the ligand solutions were determined on a Hewlett Packard 8452 diode array spectrophotometer using known extinction coefficients. All binding experiments were performed in 10 mM *K*-cacodylate (pH = 5.8), 10 mM $MgCl_2$, 10 mM KCl. Absorbance/emission wavelengths for TMR and pyronin Y were 550/574 nm and 548/562 nm, respectively. Fluorescence anisotropy was measured for a series of samples with constant amounts of input fluorophore ($[F]_0 = 30$ nM) and increasing concentrations of aptamer RNA. Dissociation constants were calculated by fitting the observed fluorescence anisotropy values to an equation of the form:

$$A = A_0 + \Delta A\{([RNA]_0 + [F]_0 + K_D) - [([RNA]_0 + [F]_0 + K_D)^2 - 4[RNA]_0[F]_0]^{1/2}\}/2$$

where A and A_0 are the observed fluorescence anisotropy values in the presence and absence of RNA, ΔA is the difference in fluorescence anisotropy observed with no and saturating concentrations of RNA, $[RNA]_0$ and $[F]_0$ are the concentrations of input RNA and fluorophore (Hamasaki *et al.*, 1998). Non-linear least squares fitting was performed using Kaleidagraph.

Accession numbers

The refined structure has been deposited with the RCSB Protein Data Bank (accession identifier 1F1T).

Acknowledgments

We thank Dr Thomas Earnest, for help in data collection at ALS. We thank Drs William Scott, Django Sussman, and Jay Nix for helpful discussions. Supported by awards to C.W. from the NIH, NSF, and David and Lucile Packard Foundation. RNA structural biology at UCSC is supported by the W.M. Keck Foundation.

References

Anderson, A. C., O'Neil, R. H., Filman, D. J. & Frederick, C. A. (1999). Crystal structure of a brominated RNA helix with four mismatched base-pairs: an investigation into RNA conformational variability. *Biochemistry,* **38**, 12577-12585.

Balasubramanian, B., Pogozelski, W. K. & Tullius, T. D. (1998). DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl Acad. Sci. USA,* **95**, 9738-9743.

Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiange, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Raed, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.

Burgstaller, P. & Famulok, M. (1994). Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angew. Chem. Int. Ed. Engl.* **33**, 1084-1087.

Carson, M. & Bugg, C. E. (1986). Algorithms for ribbons models of proteins. *J. Mol. Graph.* **4**, 121-122.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Use of chemically modified nucleotides to determine a 62-nucleotide RNA crystal structure: a survey of phosphorothioates, Br, Pt and Hg. *J. Biomol. Struct. Dynam.* **15**, 165-172.

Culver, G. M. & Noller, H. F. (1998). Directed hydroxyl radical probing of 16 S ribosomal RNA in ribosomes containing Fe(II) tethered to ribosomal protein S20. *RNA,* **4**, 1471-1480.

De La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-493.

Dieckmann, T., Suzuki, E., Nakamura, G. K. & Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA,* **2**, 628-640.

Ellington, A. D. & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature,* **346**, 818-822.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996). Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.

Feigon, J., Dieckmann, T. & Smith, F. W. (1996). Aptamer structures from A to zeta. *Chem. Biol.* **3**, 611-617.

Ferré-D'Amaré, A. R. & Doudna, J. A. (2000). Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541-556.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13-27.

Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A. & Famulok, M. (1996). RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucl. Acids Res.* **24**, 1029-1036.

Grate, D. & Wilson, C. (1999). Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl Acad. Sci. USA,* **96**, 6131-6136.

Hamasaki, K., Killian, J., Cho, J. & Rando, R. R. (1998). Minimal RNA constructs that specifically bind aminoglycoside antibiotics with high affinities. *Biochemistry,* **37**, 656-663.

Heilek, G. M. & Noller, H. F. (1996). Site-directed hydroxyl radical probing of the Rrna neighborhood of ribosomal protein S5. *Science,* **272**, 1659-1662.

Hermann, T. & Heumann, H. (1995). Determination of nucleotide distances in RNA by means of copper phenanthroline-generated hydroxyl radical cleavage pattern. *RNA,* **1**, 1009-1017.

Hermann, T. & Patel, D. J. (1999). Stitching together RNA tertiary architectures. *J. Mol. Biol.* **294**, 829-849.

Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. (1991). Conic: a fast renderer for space-filling molecules with shadows. *J. Mol. Graph.* **9**, 230-236.

Huber, A. H., Nelson, W. J. & Weis, W. I. (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell,* **90**, 871-882.

Jain, S. C. & Sobell, H. M. (1984). Visualization of drug-nucleic acid interactions at atomic resolution. VIII. Structures of two ethidium/dinucleoside monophosphate crystalline complexes containing ethidium: cytidylyl(3′-5′) guanosine. *J. Biomol. Struct. Dynam.* **1**, 1179-1194.

Jay, D. G. & Keshishian, H. (1990). Laser inactivation of fasciclin I disrupts axon adhesion of grasshopper pioneer neurons. *Nature,* **348**, 548-550.

Jay, D. G. & Sakurai, T. (1999). Chromophore-assisted laser inactivation (CALI) to elucidate cellular mechanisms of cancer. *Biochim. Biophys. Acta,* **1424**, M39-M48.

Jiang, F., Kumar, R. A., Jones, R. A. & Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature,* **382**, 183-186.

Jiang, L. & Patel, D. J. (1998). Solution structure of the tobramycin-RNA aptamer complex. *Nature Struct. Biol.* **5**, 769-774.

Jucker, F. M. & Pardi, A. (1995). GNRA tetraloops make a U-turn. *RNA,* **1**, 219-222.

Katahira, M., Moriyama, K., Kanagawa, M., Saeki, J., Kim, M. H., Nagaoka, M., Ide, M., Uesugi, S. & Kono, T. (1995). RNA quadruplex containing G and A. *Nucl. Acids Symp. Ser.* **36**, 197-198.

Kelley, S. O. & Barton, J. K. (1999). Electron transfer between bases in double helical DNA. *Science,* **283**, 375-381.

Liao, J. C., Roider, J. & Jay, D. G. (1994). Chromophore-assisted laser inactivation of proteins is mediated by the photogeneration of free radicals. *Proc. Natl Acad. Sci. USA,* **91**, 2659-2663.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3d version 2.0 - a program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.

Mosges, G., Hampel, F., Kaupp, M. & Schleyer, P. V. (1992). Experimental and theoretical studies of alkaline-earth metal coordination - X-Ray crystal structures of calcium, strontium, and barium carbazoles and *ab initio* model calculations. *J. Am. Chem. Soc.* **114**, 10880-10889.

Mueller, U., Schübel, H., Sprinzl, M. & Heinemann, U. (1999). Crystal structure of acceptor stem of tRNA(Ala) from *Escherichia coli* shows unique G.U wobble base-pair at 1.16 Å resolution. *RNA,* **5**, 670-677.

Nix, J., Sussman, D. & Wilson, C. (2000). 1.3 Å crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. *J. Mol. Biol.* **296**, 1235-1244.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, vol. 276, vol. 307-326, Academic Press, San Diego.

Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A. & Nonin, S. (1997). Structure, recognition and adaptive binding in RNA aptamer complexes. *J. Mol. Biol.* **272**, 645-664.

Sassanfar, M. & Szostak, J. W. (1993). An RNA motif that binds ATP. *Nature,* **364**, 550-553.

Scott, W. G., Finch, J. T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme. *Nucl. Acids Symp. Ser.* **81**, 214-216.

Shepard, W., Cruse, W. B. T., Fourme, R., delaFortelle, E. & Prange, T. (1998). A zipper-like duplex in DNA: the crystal structure of d(GCGAAAGCT) at 2.1 Å resolution. *Structure,* **6**, 849-861.

Steiner, G., Kuechler, E. & Barta, A. (1988). Photo-affinity labelling at the peptidyl transferase centre reveals two different positions for the A- and P-sites in domain V of 23 S rRNA. *EMBO J.* **7**, 3949-3955.

Stout, C. D., Mizuno, H., Rubin, J., Brennan, T., Rao, S. T. & Sundaralingam, M. (1976). Atomic coordinates and molecular conformation of yeast phenylalanyl tRNA. An independent investigation. *Nucl. Acids Res.* **3**, 1111-1123.

Su, L., Chen, L., Egli, M., Berger, J. M. & Rich, A. (1999). Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nature Struct. Biol.* **6**, 285-292.

Sussman, D., Nix, J. C. & Wilson, C. (2000). The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nature Struct. Biol.* **7**, 53-57.

Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D,* **55**, 849-861.

Todd, A. K., Adams, A., Powell, H. R., Wilcock, D. J., Thorpe, J. H., Lausi, A., Zanini, F., Wakelin, L. P. & Cardin, C. J. (1999). Determination by MAD-DM of the structure of the DNA duplex d[ACGTACG(5-BrU)]2 at 1.46 A and 100 K. *Acta Crystallog. sect. D,* **55**, 729-735.

Wang, F. S., Wolenski, J. S., Cheney, R. E., Mooseker, M. S. & Jay, D. G. (1996). Function of myosin-V in filopodial extension of neuronal growth cones. *Science,* **273**, 660-663.

Wild, K., Weichenrieder, O., Leonard, G. A. & Cusack, S. (1999). The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Structure Fold. Des.* **7**, 1345-1352.

Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703-730.

Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science,* **272**, 1343-1347.

Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. & Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nature Struct. Biol.* **4**, 644-649.

Edited by J. A. Doudna

(Received 27 March 2000; received in revised form 12 June 2000; accepted 12 June 2000)